

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

(11)Publication number : 11-213000

(43)Date of publication of application : 06.08.1999

(51)Int.Cl. G06F 17/30

(21)Application number : 10-017422

(71)Applicant : NIPPON TELEGR & TELEPH
CORP <NTT>

(22)Date of filing : 29.01.1998

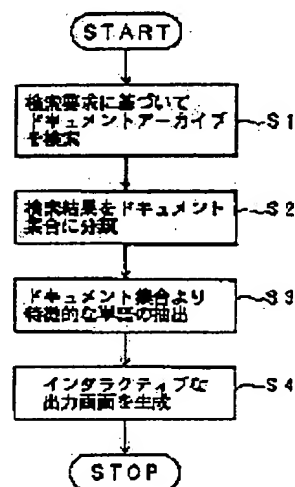
(72)Inventor : HAYASHI YOSHIHIKO

(54) INTERACTIVE INFORMATION RETRIEVAL METHOD/DEVICE AND STORAGE
MEDIUM STORING INTERACTIVE INFORMATION RETRIEVAL PROGRAM

(57)Abstract:

PROBLEM TO BE SOLVED: To execute efficient information navigation motion with the result of information retrieval as a start even if the result of information retrieval is indicated only by the list documents.

SOLUTION: In the retrieval method, information is basically retrieved based on the information request of a user with a document archive as a retrieval object and a document group in the list of the retrieval result obtained by information retrieval is classified into document sets constituted of similar documents. The list of characteristic words is extracted from the respective document sets of the retrieval result. Information of an output screen which can interactively be operated to the user based on the classified group and the list of the extracted words is generated.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-213000

(43) 公開日 平成11年(1999) 8月6日

(51) Int.Cl.⁸
G 0 6 F 17/30

識別記号

F I
G 0 6 F 15/401 3 1 0 D
3 2 0 A
15/403 3 8 0 Z

審査請求 未請求 請求項の数12 O L (全 12 頁)

(21) 出願番号 特願平10-17422

(22) 出願日 平成10年(1998) 1月29日

(71) 出願人 000004226

日本電信電話株式会社
東京都新宿区西新宿三丁目19番2号

(72) 発明者 林 良彦

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74) 代理人 弁理士 伊東 忠彦

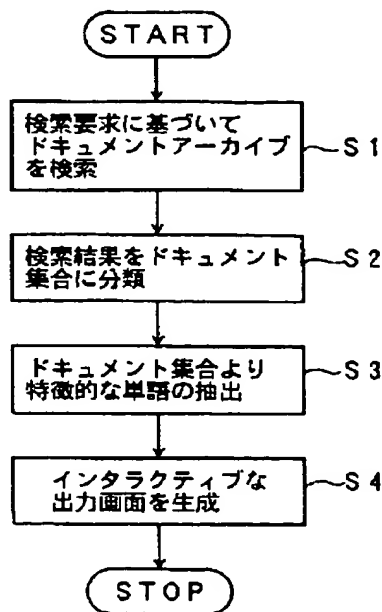
(54) 【発明の名称】 インタラクティブ情報検索方法及び装置及びインタラクティブ情報検索プログラムを格納した記憶媒体

(57) 【要約】

【課題】 情報検索の結果が単なる文書のリストによってしか提示されないために、ユーザは、情報検索の結果を起点とした効率的な情報ナビゲーション行動ができないという課題を解決した、インタラクティブなインタラクティブ情報検索方法及び装置及びインタラクティブ情報検索プログラムを格納した記憶媒体を提供する。

【解決手段】 本発明は、ドキュメントアーカイブを検索対象として、ユーザの情報要求に基づいて基本的な情報検索を行い、情報検索により取得した検索結果のリストにおけるドキュメント群を互いに類似したドキュメントにより構成されるドキュメント集合に分類し、検索結果の各ドキュメント集合から特徴的な単語のリストを抽出し、分類されたグループと、抽出された単語のリストに基づいてユーザに対してインタラクティブに操作可能な出力画面の情報を生成する。

本発明の原理を説明するための図



【特許請求の範囲】

【請求項 1】 予め構築されたドキュメントアーカイブをユーザにより入力される情報要求に従って検索するインタラクティブ情報検索方法において、前記ドキュメントアーカイブを検索対象として、前記ユーザの情報要求に基づいて基本的な情報検索を行い、前記情報検索により取得した検索結果のリストにおけるドキュメント群を互いに類似したドキュメントにより構成されるドキュメント集合に分類し、前記検索結果の各前記ドキュメント集合から特徴的な単語のリストを抽出し、分類されたグループと、抽出された単語のリストに基づいて前記ユーザに対してインタラクティブに操作可能な出力画面の情報を生成することを特徴とするインタラクティブ情報検索方法。

【請求項 2】 前記出力画面の情報を生成する際に、前記出力画面の表示と同時に、前記ユーザの次の行動を入力可能な画面情報を表示する請求項 1 記載のインタラクティブ情報検索方法。

【請求項 3】 前記ユーザにより選択された少なくとも 1 つのドキュメント集合に対して、再度分類を行い、前記検索結果を絞り込む請求項 1 記載のインタラクティブ情報検索方法。

【請求項 4】 抽出された前記単語のリストから前記ユーザにより選択された単語または、単語の組み合わせにより再度検索を実行する請求項 1 記載のインタラクティブ情報検索方法。

【請求項 5】 予め構築されたドキュメントアーカイブをユーザにより入力される情報要求に従って検索するインタラクティブ情報検索装置であって、前記ドキュメントアーカイブを検索対象として、前記ユーザの情報要求に基づいて基本的な情報検索を行う情報検索手段と、前記情報検索手段により取得した検索結果のリストにおけるドキュメント群を互いに類似したドキュメントにより構成されるドキュメント集合に分類する検索結果分類手段と、前記検索結果の各前記ドキュメント集合から特徴的な単語のリストを抽出する単語抽出手段と、前記検索結果分類手段と、前記単語抽出手段の結果に基づいて、前記ユーザに対してインタラクティブに操作可能な出力画面の情報を生成する出力情報編集手段とを有することを特徴とするインタラクティブ情報検索装置。

【請求項 6】 前記出力情報編集手段は、前記出力画面の表示と同時に、前記ユーザの次の行動を入力可能な画面情報を表示する手段を含む請求項 5 記載のインタラクティブ情報検索装置。

【請求項 7】 前記検索結果分類手段は、前記ユーザにより選択された少なくとも 1 つの前記ドキュメント集合に対して、再度分類を行う手段を含む請求

項 5 記載のインタラクティブ情報検索装置。

【請求項 8】 前記情報検索手段は、抽出された前記単語のリストから前記ユーザにより選択された単語または、単語の組み合わせにより再度検索を実行する手段を含む請求項 5 記載のインタラクティブ情報検索装置。

【請求項 9】 予め構築されたドキュメントアーカイブをユーザにより入力される情報要求に従って検索するインタラクティブ情報検索プログラムを格納した記憶媒体であって、

前記ドキュメントアーカイブを検索対象として、前記ユーザの情報要求に基づいて基本的な情報検索を行う情報検索プロセスと、

前記情報検索プロセスにより取得した検索結果のリストにおけるドキュメント群を互いに類似したドキュメントにより構成されるドキュメント集合に分類する検索結果分類プロセスと、

前記検索結果の各前記ドキュメント集合から特徴的な単語のリストを抽出する単語抽出プロセスと、

前記検索結果分類プロセスと、前記単語抽出プロセスの結果に基づいて、前記ユーザに対してインタラクティブに操作可能な出力画面の情報を生成する出力情報編集プロセスとを有することを特徴とするインタラクティブ情報検索プログラムを格納した記憶媒体。

【請求項 10】 前記出力情報編集プロセスは、前記出力画面の表示と同時に、前記ユーザの次の行動を入力可能な画面情報を表示するプロセスを含む請求項 9 記載のインタラクティブ情報検索プログラムを格納した記憶媒体。

【請求項 11】 前記検索結果分類プロセスは、前記ユーザにより選択された少なくとも 1 つの前記ドキュメント集合に対して、再度分類を行うプロセスを含む請求項 9 記載のインタラクティブ情報検索プログラムを格納した記憶媒体。

【請求項 12】 前記情報検索プロセスは、抽出された前記単語のリストから前記ユーザにより選択された単語または、単語の組み合わせにより再度検索を実行するプロセスを含む請求項 5 記載のインタラクティブ情報検索プログラムを格納した記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、インタラクティブ情報検索方法及び装置及びインタラクティブ情報検索プログラムを格納した記憶媒体に係り、特に、ドキュメントアーカイブをユーザの情報要求に従って検索する情報検索システムにおいて、検索結果を自動分類し、検索結果のドキュメント群において特徴的な単語リストを求めることにより、ユーザの次の情報検索行動を支援するための出力情報を提供し、これにより、ユーザのインタラクティブな情報検索行動を支援するためのインタラクティブ情

報検索方法及び装置及びインタラクティブ情報検索プログラムを格納した記憶媒体に関する。

【0002】

【従来の技術】従来の情報検索システムにおいては、ユーザの情報検索要求に対して検索されたドキュメントアーカイブ中の各ドキュメントについて、ユーザの情報要求に対する適合度を計算し、これによってソートされたドキュメントのリストをユーザに提示することが通常行われている。

【0003】

【発明が解決しようとする課題】しかしながら、上記の従来の情報検索システムでは、この適合度は必ずしもユーザの意向や直感を反映するものであるとは限らない。また、単なるドキュメントのリストであるため、情報検索結果を起点とするユーザの情報ナビゲーション行動を支援するには十分でない。

【0004】また、近年のインターネット上のサーチエンジンに顕著のように、膨大なドキュメント群（この場合はWWW上のページ群）を対象とする場合、リストは膨大なものになる場合も多く、リストの各要素であるドキュメントを全てユーザがチェックすることは事実上不可能と言える。本発明は、上記の点に鑑みなされたもので、情報検索の結果が単なる文書のリストによってしか提示されないために、ユーザは、情報検索の結果を起点とした効率的な情報ナビゲーション行動ができないという課題を解決する、インタラクティブなインタラクティブ情報検索方法及び装置及びインタラクティブ情報検索プログラムを格納した記憶媒体を提供することを目的とする。

【0005】

【課題を解決するための手段】図1は、本発明の原理を説明するための図である。本発明（請求項1）は、予め構築されたドキュメントアーカイブをユーザにより入力される情報要求に従って検索するインタラクティブ情報検索方法において、ドキュメントアーカイブを検索対象として、ユーザの情報要求に基づいて基本的な情報検索を行い（ステップ1）、情報検索により取得した検索結果のリストにおけるドキュメント群を互いに類似したドキュメントにより構成されるドキュメント集合に分類し（ステップ2）、検索結果の各ドキュメント集合から特徴的な単語のリストを抽出し（ステップ3）、分類されたグループと、抽出された単語のリストに基づいてユーザに対してインタラクティブに操作可能な出力画面の情報

を生成する（ステップ4）。
【0006】本発明（請求項2）は、出力画面の情報を生成する際に、出力画面の表示と同時に、ユーザの次の行動を入力可能な画面情報を表示する。本発明（請求項3）は、ユーザにより選択された少なくとも1つのドキュメント集合に対して、再度分類を行い、検索結果を絞り込む。本発明（請求項4）は、抽出された単語のリス

トからユーザにより選択された単語または、単語の組み合わせにより再度検索を実行する。

【0007】図2は、本発明の原理構成図である。本発明（請求項5）は、予め構築されたドキュメントアーカイブ50をユーザにより入力される情報要求に従って検索するインタラクティブ情報検索装置であって、ドキュメントアーカイブ50を検索対象として、ユーザの情報要求に基づいて基本的な情報検索を行う情報検索手段10と、情報検索手段10により取得した検索結果のリストにおけるドキュメント群を互いに類似したドキュメントにより構成されるドキュメント集合に分類する検索結果分類手段20と、検索結果の各ドキュメント集合から特徴的な単語のリストを抽出する単語抽出手段30と、検索結果分類手段20と、単語抽出手段30の結果に基づいて、ユーザに対してインタラクティブに操作可能な出力画面の情報を生成する出力情報編集手段40とを有する。

【0008】本発明（請求項6）は、出力情報編集手段40において、出力画面の表示と同時に、ユーザの次の行動を入力可能な画面情報を表示する手段を含む。本発明（請求項7）は、検索結果分類手段20において、ユーザにより少なくとも1つ選択されたドキュメント集合に対して、再度分類を行う手段を含む。本発明（請求項8）は、情報検索手段10において、単語抽出手段30により抽出された単語のリストからユーザにより選択された単語または、単語の組み合わせにより再度検索を実行する手段を含む。

【0009】本発明（請求項9）は、予め構築されたドキュメントアーカイブをユーザにより入力される情報要求に従って検索するインタラクティブ情報検索プログラムを格納した記憶媒体であって、ドキュメントアーカイブを検索対象として、ユーザの情報要求に基づいて基本的な情報検索を行う情報検索プロセスと、情報検索プロセスにより取得した検索結果のリストにおけるドキュメント群を互いに類似したドキュメントにより構成されるドキュメント集合に分類する検索結果分類プロセスと、検索結果の各ドキュメント集合から特徴的な単語のリストを抽出する単語抽出プロセスと、検索結果分類プロセスと、単語抽出プロセスの結果に基づいて、ユーザに対してインタラクティブに操作可能な出力画面の情報を生成する出力情報編集プロセスとを有する。

【0010】本発明（請求項10）は、情報編集プロセスにおいて、出力画面の表示と同時に、ユーザの次の行動を入力可能な画面情報を表示するプロセスを含む。本発明（請求項11）は、検索結果分類プロセスにおいて、ユーザにより選択された少なくとも1つのドキュメント集合に対して、再度分類を行うプロセスを含む。

【0011】本発明（請求項12）は、情報検索プロセスにおいて、抽出された単語のリストからユーザにより選択された単語または、単語の組み合わせにより再度検

索を実行するプロセスを含む。上記のように、本発明によれば、以下のようなことが可能となり、これらは情報検索に基づく、ユーザのインタラクティブな情報ナビゲーションを支援することが可能となる。

【0012】(1) 情報検索結果が単なるリストではなく、検索結果のドキュメントがいくつかの数のグループへと自動的に分類されるため、実際にアクセスするドキュメントを決定する際の支援となる。

(2) さらに、上記で生成されたグループを1つ以上選択した結果であるドキュメント集合に対して、再度の自動分類を指示することにより、検索結果を絞り込んでいくことが可能となる。

【0013】(3) 自動分類の過程で抽出される特徴的な単語のリストを利用して、この中からユーザが適当な単語を選択して新しい検索を実行することにより、選択する単語群とその組み合わせ方によって、絞り込み的な検索や、関連するトピックに関する検索(連想型検索)が可能となる。

【0014】

【発明の実施の形態】図3は、本発明の情報検索装置の構成を示す。同図に示すシステムは、入力部100、検索サービス部200、情報検索実行部210、ドキュメントアーカイブ220、ドキュメント分類サービス部300、検索結果分類実行部310、ドキュメント特徴抽出部320、出力部400、及び出力情報編集部410から構成される。

【0015】入力部100は、ユーザ端末から送信されるシステムへの指示を受け付ける。ユーザからの指示は、

①検索のための情報要求；

②検索結果の再分類の指示；

のどちらかである。入力部100は、これらの判断を行い、①の場合には、検索サービス部200へ入力された情報要求を転送する。②の場合は、ドキュメント分類サービス部300へ入力された検索結果の再分類の指示を転送する。

【0016】検索サービス部200は、まず、入力部100から転送されてきた情報要求を情報検索実行部210へ転送する。情報検索実行部210は、予め構築されたドキュメントアーカイブ220を検索対象とした情報検索を実行し、その結果として、ユーザの情報要求に対する適合度順にソートされた、ドキュメントアーカイブ220のドキュメントのリストを検索サービス部200へ返却する。なお、情報検索実行部210に相当するテキスト検索エンジンサブシステムは、公知の技術により十分実現可能であるため、本発明ではその詳細は問わず、入力として単語の論理結合(AND結合、OR結合)を許すことと、適合度順にソートされたドキュメントリストを結果とすることのみ条件とする。

【0017】検索サービス部200は、次にドキュメン

トリストを検索結果分類実行部310へ転送し、その処理結果をさらに出力情報編集部410へと転送する。ドキュメント分類サービス部300は、転送されてきた検索結果の再分類の指示に基づいて、再分類の対象となるドキュメントの集合を求め、これを検索結果分類実行部310へ転送し、その処理結果をさらに出力情報編集部410に転送する。

【0018】検索結果分類実行部310は、検索サービス部200または、ドキュメント分類サービス部300より転送されてくるドキュメント集合(実際にはドキュメントアーカイブ220におけるドキュメントIDの集合)を入力する。検索結果分類実行部310は、まず、ドキュメント特徴抽出部320を呼び出すことにより、ドキュメントアーカイブ220における指定されたドキュメント特徴を得る。ここで、ドキュメントの特徴とは、ドキュメントに含まれる単語のリストであり、各単語には、その単語の重要性を表す重み(実数値)が付与されているものとする。即ち、ドキュメントの特徴は実数値を値とし、各要素はある単語に対応しているベクトルとして表現される。単語の重みとしては、情報検索の分野において従来より検討がなされており、単純な頻度、正規化された頻度や特徴的であるか否かを表す値(TF・IDF)などが考えられる。本発明では、単語の重みとして何を使用するかについては規定しない。また、ドキュメント特徴抽出部320では、指示されたドキュメントの特徴をその都度計算するのではなく、ドキュメントの特徴を内部データベースにキャッシュしておくなどの効率化手段が考えられるが、本発明ではその詳細は規定しない。

【0019】検索結果分類実行部310は、次に、入力されたドキュメント集合の各要素であるドキュメントに対して求められたドキュメント特徴ベクトルを総合することにより、図4に示すようなマトリクスを求める。当該マトリクスの各行はドキュメントに、各列は単語に相当する。このようなマトリクスを以下では特徴マトリクスと呼ぶ。

【0020】ここで、図4のような特徴マトリクスに基づいて、オブジェクト(ドキュメント)の自動分類を行う方法として、クラスタリングと呼ばれる手法が知られており、いくつかのアルゴリズムが提案されている(参考文献例:E.Rasmussen: Clustering Algorithms, in W.B. Frakes, R. Baeza-Yates, editors, Information Retrieval, Prentice Hall, 1992)。本発明における検索結果分類実行部310の採用するクラスタリングのアルゴリズムについては、本発明では規定しないが、入力された特徴マトリクスとグループの個数に対して、図5に示すようなマトリクス(分類結果マトリクスと呼ぶ)で表される分類結果が得られるものとする。

【0021】また、図6に示すように、各グループにおいて特徴的な単語のリストも求めるものとする。このよ

うな単語のリストは、分類結果マトリクスの各行から、大きな重みを持つ単語を抽出することで、容易に構成することができる。ここで、リストの大きさ（単語の数）を陽に指定することもできるし、ある一定の値以上の重みを持つ単語のみを対象としてリストを構成してもよいが、本発明では、その方法については規定しない。

【0022】出力情報編集部410は、検索結果分類実行部310から以下のデータ

(1) グループに属するドキュメントのリスト；

(2) グループを特徴付ける単語のリスト（単語＋重み）；

を受け取り、ユーザによるインタラクティブな情報検索行動を支援するための出力画面（同時に次の行動を入力する画面でもある）情報を生成する。

【0023】出力部400は、出力情報編集部410から転送されてきた画面情報をユーザの端末へと転送する。

【0024】

【実施例】以下、図面と共に本発明の実施例を説明する。以下の実施例では、本発明のインタラクティブ情報検索システムをWWW（World Wide Web）上のサーチエンジンに適用した場合の具体例を説明する。図7は、本発明の一実施例の検索要求入力画面の例を示しており、ユーザ端末に表示される初期画面の例である。この画面例においては、まず、情報要求を表すフレーズを入力する領域が最上部に設定されており、ユーザは、“オリンピックでのボランティア”なる情報要求を入力したものとする。第2の領域では、自動分類処理に関する設定を行えるようになっており、この例では、分類グループの数は自動的に決定すること、結果は、フレームを利用して表示することを指示している。第3の領域では、テキスト情報検索に関する基本的な設定が行えるようになっており、この例では、情報要求フレーズから抽出した語をAND条件で結んで検索を行うこと、また、検索結果のドキュメント件数を10件に制限することなどを指示している。以下では、この入力例に即して説明を行う。

【0025】入力部100は、上記のような入力画面を通してユーザ端末から送信されてくる要求のタイプに従って、検索サービス部200、または、ドキュメント分類サービス部300へユーザの要求を転送する。上記の例の場合、情報検索のための情報要求であるため、検索サービス部200へと要求を転送する。検索サービス部200は、転送されてきた情報要求から検索条件に用いる単語を抽出する。この例において転送されてくる情報要求は、“オリンピックのボランティア”であり、このフレーズからは、“オリンピック”と“ボランティア”の2語を抽出する。なお、自然言語表現から単語を抽出する処理は、既存の形態素解析技術などを適用することにより、容易に実現可能であるので、本発明では、その詳細については規定しない。上記の情報要求及び検

索条件からは、

（AND “オリンピック” “ボランティア”）

なる検索式が生成され、この検索式は、情報検索実行部210へ転送される。

【0026】図8は、本発明の一実施例の情報検索結果のドキュメントリストの例を示し、上記の検索式によって情報検索実行部210がドキュメントアーカイブ220を検索対象として検索を行った結果例である。情報検索実行部210に適用されるような通常のテキスト検索サブシステムは、ここに示されたような情報以外の情報も返却することもできる（実際、後で示される画面例では、そのような情報も利用した表示が行われている）が、図8では、以下の説明に必要な最小限の情報のみを示す。即ち、情報検索結果の各ドキュメントに対しては、そのドキュメントアーカイブ220内におけるドキュメントID（この例では4桁の正数）、情報要求との適合度を表すスコア（値が大きいほど情報要求に適合している）、ドキュメントのタイトル（これは、説明を分かりやすくするために導入した）が返却されるものとする。

【0027】検索サービス部200は、図8に示されるような情報検索実行部210による検索結果を検索結果分類実行部310へと転送する。検索結果分類実行部310は、まず、ドキュメント特徴抽出部320を呼び出すことにより、転送されてきたドキュメントの特徴を得る。図9は、本発明の一実施例のドキュメント特徴ベクトルの例であり、検索結果リストにおける第1要素の例である。同図は、前述の図8の検索結果リストにおける第1要素である（ID=1000、タイトル=平成9年度予算の主要事業）なるドキュメントの特徴ベクトルの概念を説明するための図である。説明の簡単化のため、このドキュメントは5つの検索対象単語（T1-T5）のみを示す。

【0028】検索結果分類実行部310が、次に、入力されたドキュメント集合の各要素であるドキュメントに対して求められたドキュメント特徴ベクトルを総合することにより、図4に示すようなマトリクスを求める。図10は、本発明の一実施例の特徴マトリクスの例であり、検索結果リストに対する特徴マトリクスを説明するための図である。同図は、前述の図8の検索結果リストに対する特徴マトリクスを示しており、同図では、図9と同様に説明の簡単化のため、検索結果の10のドキュメントはいずれも5つの検索対象単語から構成されるものとするが、実際においては、検索対象単語の数はこれにとどまるものではない。

【0029】検索結果分類実行部310は、次に図10の特徴マトリクスに対してクラスタリングアルゴリズムを実行する。前述したように、いくつかのクラスタリングのアルゴリズムが提案されているので、本発明における検索結果分類実行部310は、適当なアルゴリズムを実装していると仮定する。図11は、本発明の一実施例

の分類結果マトリクスの例を示す。同図は、図10の特徴マトリクスに対してクラスタリングアルゴリズムを実行した結果の分類結果マトリクスを説明するものである。図11に示すように、この例においては、10のドキュメントが2つのグループ（1つは6つのドキュメントからなり、もう1つは4つのドキュメントからなる）へ自動分類されたものである。

【0030】図12は、本発明の一実施例の特徴的単語リストの例である。同図は検索結果分類実行部310の処理結果として、分類結果マトリクスと同時に得られる特徴的な単語リストを説明するための図である。同図において、第1のドキュメントグループにおける特徴的な単語がその重みと共に示されている（ここでは、重みが3以上の単語のみを表示した）。

【0031】上記のような検索結果分類実行部310の結果は、出力情報編集部410へと転送される。出力情報編集部410は、転送されてきたデータに基づいて、ユーザによるインタラクティブな情報検索行動を支援するための出力画面（同時に次の行動を入力する画面でもある）情報を生成する。出力部400は、出力情報編集部410から転送されてきた画面情報をユーザの端末へと転送する。

【0032】図13は、本発明の一実施例の出力画面の例を示す。同図は、出力部400によりユーザの端末に転送される具体的な出力画面の例である。同図に明らかなように、画面は、①左上部、②右上部、③下部の3つの領域に分割されている。

①左上部の領域には、自動分類処理の結果として生成されたドキュメントグループ（画面ではクラスターと記されている）の情報が表示される。図13においては、6ドキュメント（画面では6 Pages と記されている）からなる「クラスター1」と、4ドキュメントからなる「クラスター2」が生成されたことが示されている。また、それぞれのドキュメントグループ（クラスター）における特徴的な単語のリストが、ドキュメントグループの特性をユーザに示すものとして表示されている。

【0033】②右上部の領域には、生成されたドキュメントグループの数のサブ領域から構成され、各サブ領域は、一つのドキュメントグループに属するドキュメントの情報が表示される。図13の例においては、各ドキュメントに対して、タイトル、適合度（score 181）などと表示されている。他に様々な付加情報が表示されているが、これらは本発明の説明には関係しないので、説明は省略する。

【0034】③下部の領域には、各ドキュメントグループから抽出された特徴単語群を総合することにより構成した単語のリストが表示される。このような画面構成において特徴的な点は、左上部の領域と下部の領域が単なる情報表示だけでなく、ユーザの次の行動を支援する

（次の行動を入力する領域にもなっている）ように構成

されている点である。即ち、

（A） 左上部の領域においては、チェックボックスをチェックすることにより、一つ以上のドキュメントグループを選択することによって、再分類を行わせることができる。図13の例においては顕著でないが、検索結果件数の上限をもっと多くした場合や多数のドキュメントグループが生成された場合など、適切な数のグループを選択して再分類を行わせることにより、検索結果の特性をさらに詳しく調べることが可能となる。これは、必要・有用な情報へユーザを効率よく導く情報ナビゲーション支援となり得る。

【0035】（B） 下部の領域においては、システムから提示された単語のいくつかをチェックボックスをチェックすることにより選択し、また、他に加えたい単語や表現があれば、それらを追加キーワードのフィールドに入力し、単語を結合する検索条件（AND/OR）を選択することにより、新たな検索を実行することができる。このことは、検索結果の自動分類を行う過程により抽出された特徴的な単語を次の検索に利用できることを示しており、前回の検索をさらに絞り込むような検索や、直接には関係しないが、緩く関連した単語を用いることによる連想的な検索などが可能となる。これは、必要・有用な情報へのユーザを効率よく導く情報ナビゲーション支援となり得る。

【0036】以下に、（A）、（B）の場合の具体的な画面例を示す。図14は、本発明の一実施例の再分類の指示例を示す。同図は、図13における（A）左上部の領域において、6つのドキュメントからなる「クラスター1」を選択した状況を示している。この状態で、“Go!”と記されたボタンを押すことにより、“クラスター1”をさらに再分類させることができる。

【0037】図15は、本発明の一実施例の再分類結果の表示例である。同図は、再分類の結果、表示される画面例である。同図においては、6つのドキュメントが5つのドキュメントからなるグループと1つのドキュメントからなるグループの2つに分類されたことを示している。図15を図13と比較すると明らかなように、双方の画面は同じ構造を持っている。このことは、再分類の結果をさらに再分類させたり、再分類の処理において抽出された特徴的単語を利用して新たに検索実行することが可能なことを示す。

【0038】図16は、本発明の一実施例の再検索結果の指示例を示す。同図は、図13における（B）下部の領域に表示されている特徴的な単語のリストから“オリンピック”、“対策”、“環境”、“長野”の4単語を選択し、これらをAND結合することを指定した状況を示している。この状態で、“Search”と記されたボタンを押すことにより、上記の条件による新しい検索を実行させることができる。

【0039】図17は、本発明の一実施例の再検索結果

の表示例を示す。同図は、上述の新しい条件による検索の結果、表示される画面例である。この条件下では、前回と同様に10のドキュメントが検索されているが、今度は、8つのドキュメントからなるグループと2つのドキュメントからなるグループが生成されている。また、当然ながら図17の画面は、図13の画面と同様の構造を持っており、ドキュメントグループの再分類や提示された単語を選択することによる新たな検索の実行が可能なることを示す。

【0040】また、上記の実施例では、図3に示す構成に基づいて情報検索について説明したが、この例に限定されることなく、上記の処理をプログラムとして構築し、検索を行うコンピュータに接続されるディスク装置や、フロッピーディスク、CD-ROM等の可搬記憶媒体に格納しておき、本発明を実行する際に適宜インストールすることにより、本発明を容易に実現することが可能となる。

【0041】なお、本発明は上記の実施例に限定されることなく、特許請求の範囲内で種々変更・応用が可能である。

【0042】

【発明の効果】上述のように、本発明によれば、情報検索結果の自動分類や、検索結果の部分集合に対する再自動分類による検索結果の構造化、自動分類の過程で抽出された特徴的な単語を組み合わせることによる次の段階の検索支援が可能となり、これらは、情報検索に基づくユーザの情報ナビゲーションを支援する。

【図面の簡単な説明】

【図1】本発明の原理を説明するための図である。

【図2】本発明の原理構成図である。

【図3】本発明の情報検索装置の構成図である。

【図4】本発明の特徴マトリクス概念図である。

【図5】本発明の分類結果マトリクス概念図である。

【図6】本発明の特徴単語リスト概念図である。

【図7】本発明の一実施例の検索要求入力画面の例である。

【図8】本発明の一実施例の情報検索結果のドキュメントリストの例である。

【図9】本発明の一実施例のドキュメント特徴ベクトルの例である。

【図10】本発明の一実施例の特徴マトリクスの例である。

【図11】本発明の一実施例の分類結果マトリクスの例である。

【図12】本発明の一実施例の特徴的単語リストの例（重み3以上を残す場合の例）である。

【図13】本発明の一実施例の出力画面例である。

【図14】本発明の一実施例の再分類の指示例である。

【図15】本発明の一実施例の再分類結果の表示例である。

【図16】本発明の一実施例の再検索の指示例である。

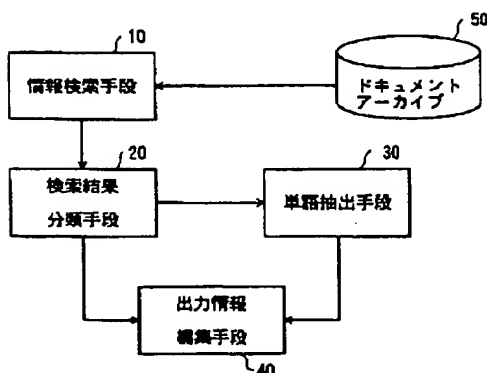
【図17】本発明の一実施例の再検索結果の表示例である。

【符号の説明】

- 10 情報検索手段
- 20 検索結果分類手段
- 30 単語抽出手段
- 40 出力情報編集手段
- 50 ドキュメントアーカイブ
- 100 入力部
- 200 検索サービス部
- 210 情報検索実行部
- 220 ドキュメントアーカイブ
- 300 ドキュメント分類サービス部
- 310 検索結果分類実行部
- 320 ドキュメント特徴抽出部
- 400 出力部
- 410 出力情報編集部

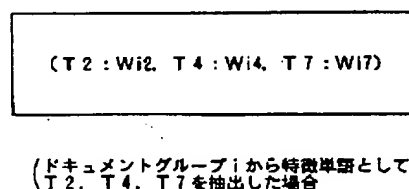
【図2】

本発明の原理構成図



【図6】

本発明の特徴単語リストの概念図



【図14】

本発明の一実施例の再分類の指示例

2個のクラスターを1秒で作りました

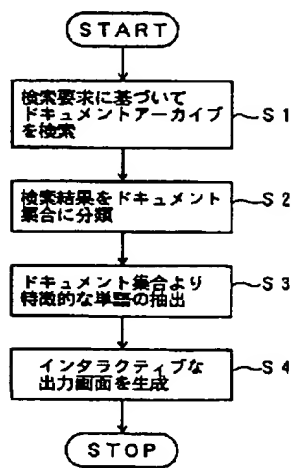
☒ クラスター 1 (6 Pages)
 (ガイド:11.2 新調:7.1対策:7.1大会
 :7.1年度:6.5平成:6.4カッ
 プ:5.5整備:5.4補正:5.4研修:5.4)

☐ クラスター 2 (4 Pages)
 (ガイド:5.8環境:5.6長野:5.0
 オリンピック:4.5保全:3.6技
 術:3.3農業:3.1センター:2.9社
 会:3.8地域:2.7)

再クラスタリングしますか?

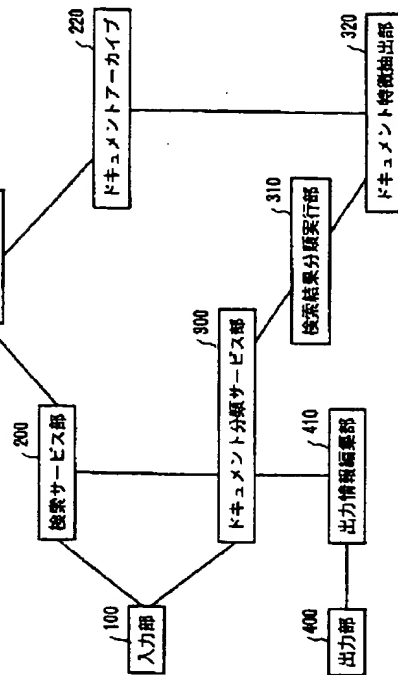
【図1】

本発明の原理を説明するための図



【図3】

本発明の情報検索装置の構成図



【図4】

本発明の特徴マトリクスの概図

	T1	T2	T3	Tn
D1	W11	W12	W13	W1n
D2	W21	W22	W23	W2n
Dm	Wm1	Wm2	Wm3	Wmn

(D_i:ドキュメント_i, T_j:単語_jの重み)
(W_{ij}:ドキュメント_iにおける単語_jの重み)

【図10】

【図5】

本発明の分類結果マトリクスの概図

	T1	T2	T3	Tn
G1: (D _a , D _b , ...)	W11	W12	W13	W1n
G2: (D _a , D _b , ...)	W21	W22	W23	W2n
Gg: (D _p , D _q , ...)	Wg1	Wg2	Wg3	Wgn

(G_i:ドキュメントグループ_i, グループ_iに属するドキュメント集合,
(T_j:単語_j, W_{ij}:ドキュメントグループ_iにおける単語_jの重み)

【図9】

本発明の一実施例のドキュメント特徴ベクトルの例

(T1(特徴):5, T2(数):5, T3(オリンピック):1, T4:(大会):0, T5(長野):0, ..., Tn(オランダ):3)

	T1 (特徴)	T2 (数)	T3 (オリンピック)	T4 (大会)	T5 (長野)
D1 (ID=1000)	5	5	1	0	0
D2 (ID=2000)	4	2	1	1	2
D3 (ID=3000)	1	1	1	0	0
D4 (ID=4000)	3	5	1	2	0
D5 (ID=5000)	0	0	8	8	20
D6 (ID=6000)	1	1	1	0	0
D7 (ID=7000)	0	0	8	8	20
D8 (ID=8000)	2	0	7	4	8
D9 (ID=9000)	1	0	10	8	18
D10 (ID=9999)	0	1	1	0	0

【図 7】

本発明の一実施例の検索要求入力画面の例

フレーズを入力して下さい

オリンピックでのボランティア

検索実行

リセット

自動分類の設定

分類グループのみ

結果のフレーム表示

自動決定

◎する ○しない

検索の基本条件

AND/OR条件

項目指定

キーワード選択

検索件数 (の上限)

○通常モード (OR)
◎絞り込みモード (AND)

☐ タイトル
☐ アンカー

◎英訳なし
○入力+英訳
○英訳のみ

10

【図 1 2】

本発明の一実施例の特徴的単語リストの例
(重み 3 以上を残す場合の例)

(T2(重):15, T1:(対象):14, T3(オリンピック):8, T4(金):3)

【図 8】

本発明の一実施例の情報検索結果のドキュメントリストの例

ID	Score	Title
1000	181	平成 9 年度予算の主要事業
2000	181	長野オリンピック冬季競技大会ボランティア活動報告
3000	128	ビルマ：イベント：ビルマ建国記念祭
4000	128	2002ワールドカップインタビュー
5000	128	NTT-Naganoオリンピック ウェブ
6000	128	最新情報
7000	128	NTT-Naganoオリンピック ウェブ
8000	123	エコロジー カフェ-J
9000	128	長野、Japan / 長野なんでも情報の Information
9999	128	会社案内と採用情報

本発明の一実施例の再検索結果の表示例

- 12 -